Aula 3

MEDIDAS DE DISPERSÃO, ASSIMETRIA E CURTOSE.

META

Mensurar a dimensão média do afastamento dos valores de um conjunto de dados em relação à determinada Medida de Tendência Central, bem como de que maneira estes dados estão alocados em torno desta Medida.

OBJETIVOS

Ao final desta aula, o estudante deverá:

Calcular parâmetros de Dispersão, dando ênfase aos de uso mais geral como: a variância, o desvio padrão e o coeficiente de variação. Você deverá entender o significado destas medidas para dados agrupados em Rol, bem como para aqueles organizados em distribuições de freqüências por valores ou intervalos de valores.

Mensurar o comportamento dos dados em torno de um parâmetro médio, sendo a média aritmética a mais indicada. Com este cálculo você deverá saber utilizar Medidas que avaliam não só a Dispersão, bem como Assimetria e Curtose.

PRÉ-REQUISITO

Todo o conhecimento relativo ao uso e calculo de Medidas de Tendência Central, além de ter em mãos: Papel, Calculadora ou Computador para realização dos cálculos.

INTRODUÇÃO

Olá! Tudo bem? Vamos dar mais um importante passo dentro da Estatística Descritiva? Para isto é muito importante o domínio do conceito, utilidade e metodologia de calculo das medidas de Tendência Central. Sem este conhecimento é impossível entender os temas a serem tratados nesta aula. Sempre que sentir necessidade recorra ao que foi tratado na aula anterior para uma melhor compreensão do que vai ser estudado.

Nesta aula você vai continuar trabalhando com a Estatística Descritiva, sendo que agora vai aprender o significado de Dispersão. Com esta medida vai poder mensurar o afastamento médio entre todos os valores de uma série e sua medida de tendência central. Dispondo desta informação vai poder classificar qualquer conjunto de dados quanto à homogeneidade e heterogeneidade e trabalhar de forma precisa esta característica nas populações ou amostras que venha a investigar.

Além da dispersão vai tomar conhecimento de outras Medidas Estatísticas muito importantes quando se investiga qualquer conjunto de dados. Uma delas é a Assimetria, com ela vai poder comparar o comportamento de conjuntos de dados com um modelo de perfeita distribuição de valores em torno da média aritmética que são os conjuntos simétricos, mais conhecidos por Distribuição Normal. Este conhecimento também mostra que quanto menor a Dispersão maior a probabilidade de termos dados com baixa Assimetria e consequentemente melhor será o conjunto observado. Finalmente vai entender o que é Curtose para avaliar a altura da maior ordenada de uma distribuição simétrica.

A aula ora apresentada servirá de guia para orientá-lo na realização destes cálculos, cabe a você fazer uso destas ferramentas e fazer mais esta aula.

MEDIDAS DE DISPERSÃO

Dispersão - É o afastamento de todos os valores de uma série em relação a média aritmética ou mediana. De acordo com a grandeza dos afastamentos as séries estatísticas podem ser: homogêneas (fraca dispersão quando $C_v < 30\%$) e heterogêneas (forte dispersão quando $C_v \geq 30\%$)

x_i: 5, 5, 5, 5, 5 $\Rightarrow$ dispersão nula
y_i: 3, 4, 5, 6, 7 $\Rightarrow$ fraca dispersão
z_i: 1, 2, 5, 8, 9 $\Rightarrow$ forte dispersão

Analisando as distribuições acima verificamos que todas possuem a mesma média aritmética, contudo revelam diferenças entre elas, sendo a principal, a que se refere ao grau de concentração de valores, em torno de um valor médio, que é mensurado pela dispersão absoluta ou relativa.

Dispersão Absoluta: dimensiona o afastamento médio entre os termos de uma série e seu valor médio. Seu resultado é expresso na mesma unidade de medida dos dados pesquisados.

Principais Medidas de Dispersão Absoluta

Variância - é a média quadrática dos desvios tomados em relação à média aritmética. Para dados não agrupados em distribuição de freqüências a variância é obtida por:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} \text{ (população)} \quad \text{e} \quad s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} \text{ (amostras)}$$

Para dados agrupados em distribuição de freqüências a variância é obtida pelas seguintes expressões:

Variável: $x_i = x_1, x_2 ... x_n$ Freqüências: $f_i = f_1, f_2 ... f_n \Rightarrow \sum f_i = n$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i} \quad \text{ou} \quad \sigma^2 = h^2 \left[\frac{\sum f_i \alpha_i^2}{\sum f_i} - \left(\frac{\sum f_i \alpha_i}{\sum f_i} \right)^2 \right]$$

No caso de amostras a variância é representada por s^2 e o denominador da expressão por (n-1) no lugar de n ou do $\sum f_i$.

Desvio Padrão (σ ou s) $\Rightarrow$ É a raiz quadrada da variância. Devemos calcular o desvio padrão quando procuramos um parâmetro que se revista de um máximo de estabilidade, além de uma plena utilização no cálculo de outras estatísticas.

Principais Propriedades do Desvio Padrão e da Variância

1. O desvio padrão ou a variância de uma constante é zero.

$x_i = a, a, a \ldots \Rightarrow \bar{x} = a \Rightarrow$ Dispersão nula.

2. Somando ou subtraindo uma constante a todos os valores da série, o desvio padrão e a variância não se alteram, enquanto a média aritmética ficará aumentada ou diminuída, respectivamente, desta constante.

$x_i = x_1, x_2, \ldots x_n$ Média = $\bar{x}$

$y_i = (x_1 + a), (x_2 + a), \ldots (x_n + a)$ Média = $\bar{x} + a$

Desvios de y_i em torno da média

$(x_1 + a) - (\bar{x} + a) = x_1 - \bar{x}$

$(x_2 + a) - (\bar{x} + a) = x_2 - \bar{x}$

$\vdots$

$(x_n + a) - (\bar{x} + a) = x_n - \bar{x}$

Variância e desvio padrão igual a série x_i

3. Multiplicando ou dividindo os valores da série por uma constante, o desvio padrão e a média aritmética ficarão multiplicados ou divididos, respectivamente, pela constante. A variância ficará multiplicada ou dividida pelo quadrado da constante.

Desvio de y_i em torno da média aritmética

$x_i = x_1, x_2, \ldots, x_n$ Média = $\bar{x}$

$y_i = x_1a, x_2a, \ldots, x_na$ Média = $\bar{x}$ a

$x_1a - a = a(x_1 - \bar{x})$

$x_2a - a = a(x_2 - \bar{x})$

$\vdots$

$x_na - a\bar{x} = a(x_n - \bar{x})$

4. A variância combinada de duas ou mais séries de valores é calculada pela seguinte expressão: Para dados populacionais (n -1) é substituído por n.

$$s^2 = \frac{(n_1 - 1)[s^2_1 + (\bar{x} - \bar{x}_1)^2] + (n_2 - 1)[s^2_2 + (\bar{x} - \bar{x}_2)^2] + ... + (n_k - 1)[s^2_k + (\bar{x} - \bar{x}_k)^2]}{(n_1 + n_2 + ... n_k) - k}$$

$$s^2 = \frac{S(n_i - 1)[s_i^2 + (\bar{x} - \bar{x}_i)^2]}{n_i - k}$$

DISPERSÃO RELATIVA

Mede a relação entre a dispersão absoluta e o valor médio da série, com resultado expresso em porcentagens.

Tem grande aplicação, visto que possibilita comparar dispersão de universos diferentes (em unidade de medida e/ou valor médio), já que seus resultados são expressos em percentagens devido à relação entre as medidas de dispersão absoluta e de tendência central. São medidas usadas para análises comparativas.

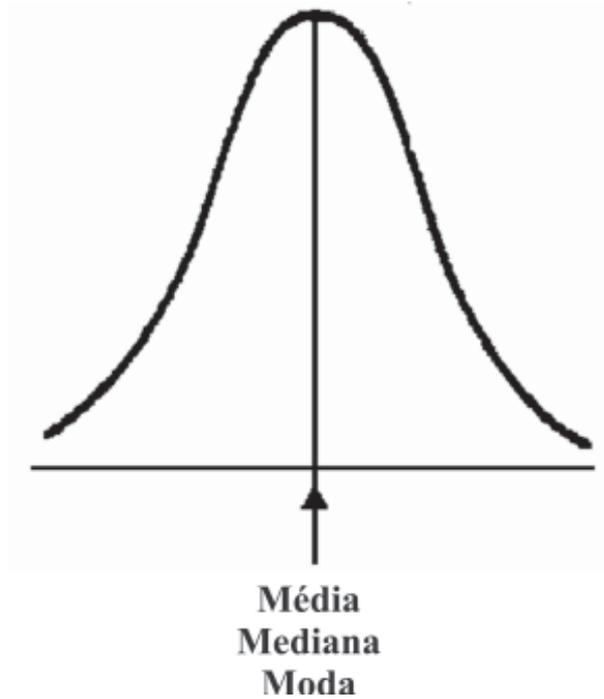
Coeficiente de variação de *PEARSON* - é a relação percentual entre o desvio padrão e a média aritmética. $CV = \frac{s}{\bar{x}} . 100$

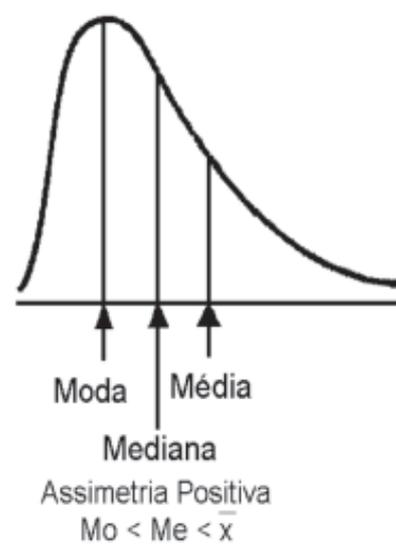
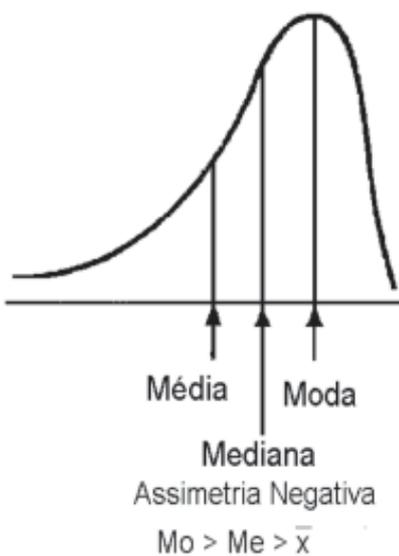
Coeficiente de Variação de *THORUDIKE*: $Cv_t = \frac{s}{M_e} . 100$

ASSIMETRIA

É o grau de desvio ou afastamento da simetria de uma distribuição. Quando a curva é simétrica, a média, a mediana e a moda coincidem, num mesmo ponto, de ordenada máxima, havendo um perfeito equilíbrio na distribuição. Quando o equilíbrio não acontece, isto é, a média, a mediana e a moda recaem em pontos diferentes da distribuição esta será assimétrica; enviesada a direita ou esquerda.

Distribuição Simétrica ⇒ Assimetria (S) = 0

$M_o = M_e = \bar{x}$ e $Q_3 - M_e = M_e - Q_1$

Distribuição assimétrica ***Negativa*** ou enviesada a esquerda - quando os valores se concentram na extremidade superior da escala e se distribuem gradativamente em direção à extremidade inferior.

Distribuição assimétrica ***Positiva*** ou enviesada a direita quando os valores se concentram na extremidade inferior da escala e se distribuem gradativamente em direção à extremidade superior.

Cálculo Da Assimetria: critério de **Bowley**

$Q_3 - M_e \neq M_e - Q_1 \quad \Rightarrow \quad Q_3 + Q_1 - 2M_e \neq 0$ (assimetria absoluta)

$$S = \frac{Q_3 + Q_1 - 2M_e}{Q_3 - Q_1}$$ (assimetria relativa); sendo: $-1 \leq S \leq 1$

Será ***Positivo*** a medida que o terceiro quartil se afasta da mediana, enquanto que o primeiro quartil se aproxima da mesma, tendo como limite: $Q_1 = Q_2$, quando a assimetria assume o valor máximo positivo: (S = 1).

$$S = \frac{Q_3 - Q_1}{Q_3 - Q_1} = 1$$

Será ***Negativo*** a medida que o primeiro quartil afasta-se da mediana, enquanto o terceiro quartil aproxima-se da mesma, dando como limite: $Q_3 = Q_2$, quando a assimetria assume valor máximo negativo:

$$S = \frac{-Q_3 + Q_1}{Q_3 - Q_1} = -1$$

Critério de ***Kelley*** **-** Para corrigir parte do inconveniente de se desprezar 50% das ocorrências (Critério de *Bowley*), *Kelley* aconselha o uso dos "**Centis**" eqüidistantes da mediana, tais como C_{10} e $C_{90,}$ para cálculo da assimetria.

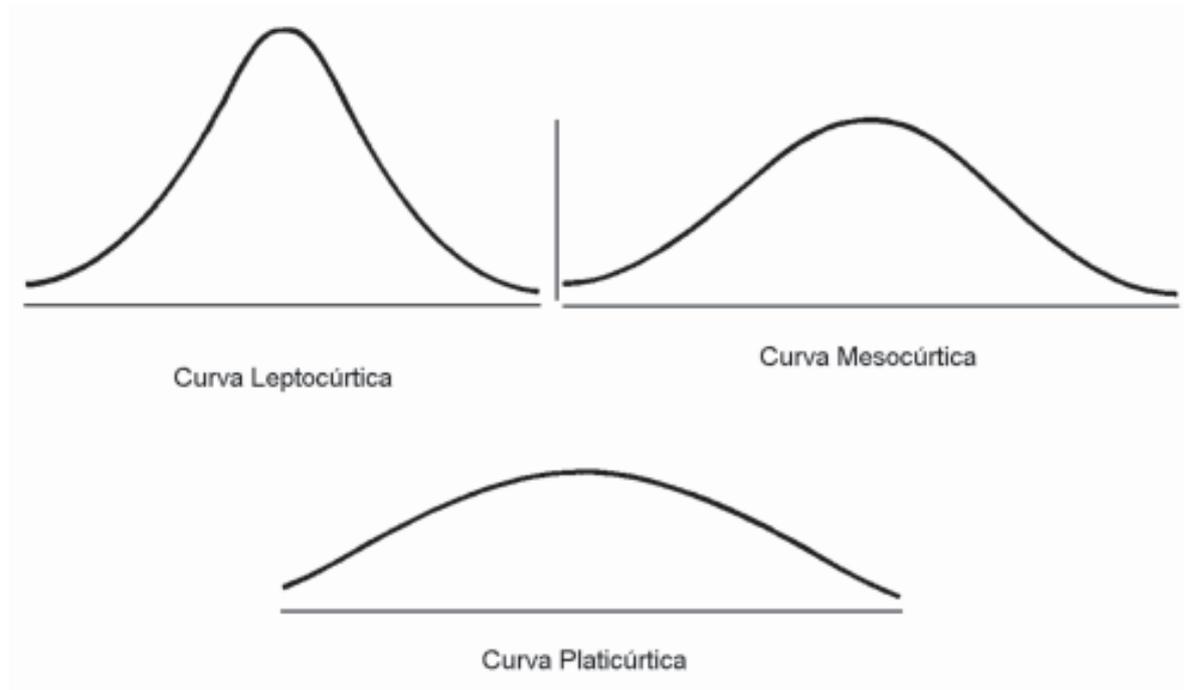
$$S = \frac{C_{90} + C_{10} - 2M_e}{C_{90} - C_{10}}$$ Cujos limites variam também de [-1 a +1]

Coeficiente de Assimetria de ***Pearson*** – *Á medi*da que a distribuição deixa de ser simétrica, a média, a mediana e a moda vão se afastando, aumentando cada vez mais a diferença entre elas.

$$\alpha = \frac{3(\bar{x} - Me)}{s} \quad \text{ou} \quad \alpha = \frac{(\bar{x} - Me)}{s}$$

CURTOSE

É o grau de achatamento de uma distribuição, em relação a distribuição normal. A curtose pode ser de três tipos:

Mesocúrtica - quando a distribuição é normal.
Leptocúrtica - quando a distribuição é mais pontiaguda que a normal
Platicúrtica - quando a distribuição é mais achatada que a normal.

MEDIDAS DE CURTOSE

Coeficiente Percentílico de Curtose:

$$C = \frac{D_q}{C_{90} - C_{10}} \quad \Rightarrow \quad C = \frac{Q_3 - Q_1}{2(C_{90} - C_{10})}$$

Se C = 0,263 a distribuição é Mesocúrtica
Se C < 0,263 a distribuição é Leptocúrtica
Se C > 0,263 a distribuição é Platicúrtica

EXERCÍCIO RESOLVIDO

1. Determinar a variância das amostras constituída dos seguintes elementos.

x_i = 7; 10; 12; 15; 16; 18; 20 e y_i = 12; 15; 18; 23; 25; 26; 28.

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

$$s^2 = \frac{49 + 16 + 4 + 1 + 4 + 16 + 36}{6} \qquad s^2 = \frac{81 + 36 + 9 + 4 + 16 + 25 + 49}{6}$$

$s^2 = 16$ e $s = 4$ $\qquad s^2 = 36{,}6667 \Rightarrow s = 6{,}06$

2. Calcular a variância da questão 1.06

$$s^2 = h^2\left[\frac{\sum f_i\alpha_i^2}{\sum f_i - 1} - \left(\frac{\sum f_i\alpha_i}{\sum f_i - 1}\right)^2\right] = 25\left[\frac{630}{349} - \left(\frac{250}{349}\right)^2\right]$$

$$s^2 = 25 \cdot 1{,}2921 \Rightarrow s^2 = 32{,}3025 \Rightarrow s = 5{,}68$$

ATIVIDADES

1.Um pesquisador da rádio XY aborda 40 transeuntes ao acaso e pergunta-lhes a idade obtendo as seguintes informações: Apresente os dados na forma de uma distribuição de freqüências, por classes e calcule a média e o desvio padrão.

38	21	16	32	39	35	26	39	25	39	22	23	27	44	30	32	35	28	17	26
22	28	39	18	37	42	40	39	22	21	40	21	15	26	43	40	37	33	24	30

2. Foram encontrados os seguintes valores para a transaminase da alamina em indivíduos normais: Calcule a média e o desvio padrão dos dados: 12; 15; 18; 12; 17; 17; 16; 16; 14; 12; 14; 17; 16; 15; 16; 18; 16; 12; 18; 18.

3. Cronometrando o tempo para várias provas de uma gincana automobilística, encontramos para duas das equipes participantes:
- Equipe I (40 provas): tempo médio = 45s e variância = $400s^2$
- Para a Equipe II, compomos o seguinte quadro:

Tempo(s)	20	40	50	80	Total
Nº de provas	10	15	30	5	60

Qual a equipe que apresentou resultado mais homogêneo?

4. Uma prova consta de três questões com pesos iguais a 3,2 e 1, respectivamente, para a primeira, a segunda e terceira questão. Se um aluno obteve 8,0 na prova 8,5 na primeira questão e 6,5 na segunda, que grau ele conseguiu na terceira questão?

5. Os dados abaixo são comprimento de crianças ao nascer. Calcular: média aritmética, variância, desvio padrão e coeficiente de Variação.
46; 47; 51; 46; 50; 50; 52; 49; 48; 51; 49; 52; 48; 47; 49.

6. Em determinada Escola foram aplicados testes a dois grupos (I e II) de alunos. Qual dos grupos possui menor dispersão.

I - 8, 7, 6, 6, 9, 10, 8, 10
II - 6, 6, 8, 9, 9, 7, 5, 6

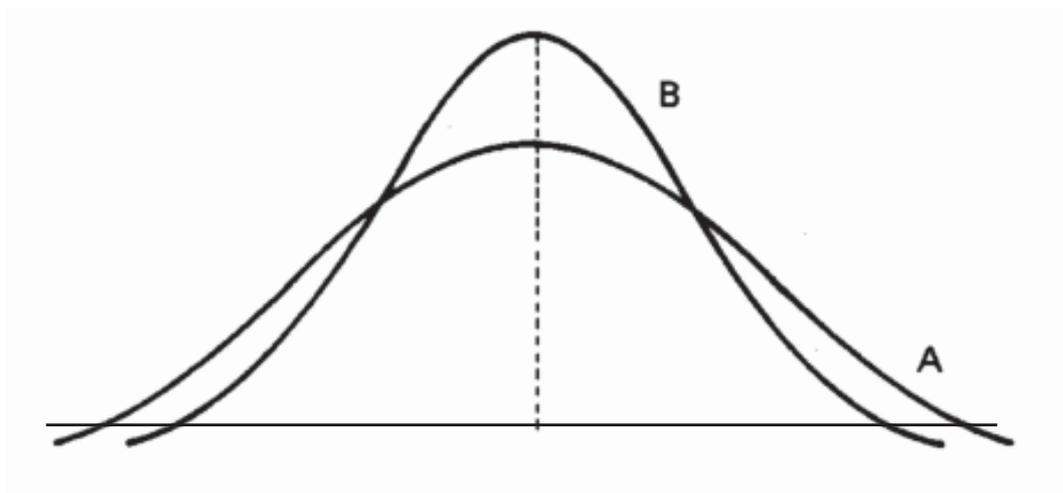
7. Uma amostra de 2.000 universitários apresentou a seguinte distribuição de alturas. Determine a média aritmética, o desvio padrão e o coeficiente de variação.

Alturas (cm)	Nº de alunos
142 |—148	46
148 |—154	88
154 |—160	332
160 |—166	676
166 |—172	614
172 |—178	158
178 |—184	65
184 |—190	21
Σ	2.000

8. O coeficiente de variação de uma distribuição é 0,083%, e o desvio é de 2,49 centímetros. Determine a média aritmética dessa distribuição.

9. Um fabricante de caixas de cartolina fabrica três tipos de caixa. Testase a resistência de cada caixa, tomando uma amostra de 100 caixas e determinando a pressão necessária para rompimento da mesma. Que tipo de caixa apresenta a menor dispersão absoluta na pressão de ruptura? Qual delas apresenta maior dispersão relativa de ruptura?

Tipos de Caixa	A	B	C
Pressão média de ruptura (bária)	190	200	300
Desvio padrão das pressões (bária)	40	50	60

10. Examinando a figura abaixo, podemos dizer:

a) ( ) O desvio padrão da distribuição A é maior do que o da distribuição B, e as médias são iguais.

b) ( ) O desvio padrão de A é menor do que o de B e as médias são diferentes.

c) ( ) O desvio padrão de A é igual ao de B, independentemente do valor da média.

d) ( ) As distribuições possuem o mesmo coeficiente de variação.

11. Realizou-se uma prova de matemática para duas turmas. Os resultados foram os seguintes:

Turma A: $\bar{x} = 5$ e $s = 2{,}5$
Turma B: $\bar{x} = 5$ e $s = 2$
Com esses resultados, podemos afirmar:
a) ( ) A turma B apresentou maior dispersão absoluta.
b) ( ) A dispersão relativa é igual à dispersão absoluta.
c) ( ) Tanto a dispersão absoluta quanto a relativa são maiores para a turma B.

d) ( ) A dispersão absoluta de A é maior do que a de B, mas em termos relativos as duas turmas não diferem quanto ao grau de dispersão das notas.

12.Suponha-se que a estrutura etária dos alunos de uma determinada unidade escolar se apresente segundo a distribuição: Calcular: média aritmética, desvio padrão, mediana, moda, assimetria e curtose.

IDADE(anos)	NÚMERO DE ALUNOS
7 |— 9	197
9 |— 11	372
11 |— 13	527
13 |— 15	114
15 |— 17	49
17 |— 19	25
19 |— 21	3
Σ	1.287

CONCLUSÃO

Com esta aula, você deve ser capaz de calcular pelas várias metodologias apresentadas qualquer medida de dispersão, assimetria e curtose. Pela dispersão deve saber o significado e a importância do afastamento entre cada valor dos dados observados e sua respectiva medida de tendência central, a partir deste conhecimento você é capaz de classificar conjuntos de números como homogêneos ou heterogêneos. Também deve saber quando aplicar uma dispersão absoluta ou relativa e como elas podem ser utilizadas para estudos e análise comparativa entre dois ou mais conjunto de dados. Será capaz de perceber que os melhores dados estatísticos são aqueles mais homogêneos, isto é: quanto menor dispersão maior é a harmonia do conjunto e melhor a qualidade deste indicador.

A assimetria deverá avaliar um outro movimento dos dados, que não aquele que pode ser observado pela dispersão. Esta medida vai levar você a perceber como os valores pesquisados estão situados em torno de sua maior ocorrência. Será capaz de concluir que quando a Assimetria de qualquer conjunto de dados for igual à zero, os valores apresentam distribuição com valores eqüidistantes em torno da principal ordenada, fazendo destes dados um conjunto de valores simétricos.

Com o conteúdo visto até esta aula, você tem boa base de parâmetros para realizar análises e interpretações dos resultados de qualquer pesquisa.

No final da aula temos uma lista de exercícios para serem resolvidos em grupos de no máximo cinco pessoas ou individual. Com certeza você vai ficar muito satisfeito com os resultados do seu desempenho.

RESUMO

Nesta aula foi apresentado o cálculo de medidas de dispersão, de assimetria e curtose. Alem do significado de cada medida e como devem ser utilizadas.

Na Dispersão, os parâmetros mais utilizados são: o desvio padrão e coeficiente de variação e são fundamentais no estudo, interpretação e análise estatística de pesquisas e outras informações que permitam o uso destes parâmetros. O desvio padrão que é resultante da raiz quadrada da variância seu valor representa o afastamento médio entre todos os valores observados e a média aritmética, lembrando que a média é uma medida de tendência central, portanto teremos afastamentos em relação a valores menores e maiores do que a média aritmética. A interpretação do desvio padrão para uma série de dados é muito importante porque ele fornece informações fundamentais para a análise de dados.

Apresentamos as propriedades do desvio padrão e de que forma este parâmetro se relaciona com os dados observados, bem como o coeficiente de variação.

Assim como no calculo da média aritmética o desvio padrão também depende da forma como os dados estão organizados, principalmente para os que se apresentam em distribuições de freqüências por intervalo de classe.

Também falamos sobre assimetria e curtose, bem como a relação destas medidas com um conjunto de dados simétricos, isto é: com uma distribuição normal.

Para analisar e interpretar qualquer conjunto de dados é preciso calcular e saber utilizar medidas de tendência central e de dispersão, além das medidas de assimetria e de curtose.

Com esta aula completamos os temas selecionados para a primeira avaliação individual.

AUTO-AVALIAÇÃO

Sou capaz de calcular medidas de dispersão como: Variância, Desvio Padrão e Coeficiente de Variação para Populações a Amostras?

Também sou capaz de calcular Assimetria e Curtose?

Sou capaz de aplicar e entender o significado das Medidas de Dispersão, Assimetria e Curtose, para conjuntos de dados investigados?

PRÓXIMA AULA

Temas relativos a experimentos aleatórios e determinísticos. Probabilidades, destacando: Espaço amostral, Eventos, Teorema da Soma e Teorema do Produto de probabilidades.

REFERÊNCIAS

RODRIGUES, PEDRO CARVALHO. **Bioestatística**. Universidade Federal Fluminense.
FONSECA, JAIRO DA. **Curso de Estatística**. Editora Atlas.
OLIVEIRA, FRANCISCO ESTEVAM MARTINS DE. **Estatística e Probabilidade**. Editora Atlas.
TANAKA. **Elementos de Estatística**. Editora McGraw.Hill.
BARBETTA, PEDRO A. **Estatística aplicada ás Ciências Sociais**. Editora da UFSC.
GÓES, LUIZ A. C. **Estatística I e II**. Editora Saraiva.
DÍAZ, FRANCISCA; LOPES, FRANCISCO JAVIER. **Bioestatística**. Editora Thomson.